How Will We Survive Disinformation?

WARNING: I'm a seer, not a scientist; I'm an entertainer, not an expert. With a sword.

[*A note about the word* will *(and not* should*) in the title. I was raised a humanist and am proud to be one, but I try to conduct my "scientific" thinking as a determinist. And that means I'm not giving advice here; I'm predicting where I think we'll end up.*]

The Source of Disinformation

These are the rules for Internet content submission.

- Anyone can put any content they want on the Internet, and those who do may keep their identities secret. (Although, if the identity is provided, or can be determined, libel and other information-related laws typically apply.)
- There shall be (effectively) no cost for putting anything on the Internet.
- There shall be no limit to the amount of information anyone or any thing can put on the internet.

It's easy to see that the Internet was designed for maximum free speech, which was the perfect design for its larval stage. But it's a mistake to leave it that way now. As each day passes it's becoming a bit more difficult to distinguish proposed truths, or ***good-faith information***, from intentional falsehoods, or ***disinformation***, as I am defining those terms here.

Curation Doesn't Work

Our current solution to the rising disinformation tide is ***curation***, but curation is failing miserably because:

1. It's self-defeating. We collect information so that we can discover the truth, but curation is the filtering out of things we know *not* to be the truth. Well, do we have the truth, or don't we? If the answer is, "we have *some* of the truth", then how are we drawing the line between the truths we have and the truths we don't? Believing that we can draw that line—which happens to be precisely where all the information worth considering shows up—is self-inflicted doublethink. Curation *can't* work—on its face.
2. It divides us into smaller and smaller filter bubbles, inside of which we drift apart from each other's truths, unable to benefit from each other's counterarguments. It's a divider of groups of people.
3. It dissolves trust, the very glue that holds societies together.
4. It surfaces extremist ideologies, and extremes are dangerous.
5. It can (and will) be overwhelmed by sheer volume. We'll no longer be able to get to the truth when we can't keep up with the work of filtering out the lies.

As is the case with boiling, *curating* an ocean is dumb, and doomed.

Filtering By Volume

Here's what's going to happen instead. We're going to stop trying to filter by content and start filtering by volume. We're going to create a subset of the Internet by charging a fee to each digital artifact that wants membership in it. Setting a price for publishing will keep the volume of disinformation manageable. We'll further demand that every digital artifact in that subset be tied to a real entity in the real world, just like money in a bank account is. If money can be tied to people, then digital artifacts can be too.

The wild and wacky Internet we have today will stay just the way it is. This new pay-to-publish subset I'm imagining will sit *inside* our Internet, and will bloom into a much more trusted (if less exciting) belief repository than the Internet at large. **This must happen because money and reputation are the only effective belief filters we have. They *will be* brought back into the mix.**

It's a Counterfeiting Problem

Think of it this way. Disinformation is *counterfeit* information. We go to great lengths to keep counterfeit *money* out of our economies because it reduces trust in the financial system. And even when the counterfeiting is done so poorly that it's easy to tell a real bill from a fake one, we still punish those who try to use the fake bills because it takes time and effort to reject them. Fake bills create a drag on the system. And worse yet, that drag makes it easier for some of the fake bills to slip through.

We face exactly the same problem with information, but we can't examine the thing that would reveal its illegitimacy—the content creator's *intent*. So, what measures can we take to tease out that intent? Charge a fee and attach a reputation. This is how *all* primate economies work. To pretend we're above that is to see the disinformation tsunami on the horizon and run to grab our surfboards so we can ride it in.

Anonymity is a Two-Sided Coin

Anonymity is great for giving the oppressed a voice, but it's also a monster enabler of disinformation. While we can continue to allow anonymity on the Internet at large, there must also be other spaces where it is strictly prohibited.

TIPT

There will be a new service (or suite of services) added to the internet to address this problem. I obviously can't know what this new Internet service will be called, but for our purposes here I will call it TIPT (The Internet Public Trust). It will be "100%" transparent, the first truly transparent system humanity builds. How will we do this? I have no idea. But it will be a system built to be maximally conspiracy-theory-proof.

Before we get started on how I imagine TIPT will work, let's quickly address three immediate questions.

Won't we be selling public opinion? Yes, but we already are.

Won't there be TIPT competitors? Yes, just like there are Wikipedia competitors. Competition is good and necessary. If there's a better way to surface trusted information than requiring money and reputation for each artifact, it should be allowed to emerge, and to win.

Won't TIPT itself be filter-bubbled? Yes, that's inevitable, but the volume and anonymity problems remain solved, and they're the main problems. In other words, any subset of TIPT will still contain only information that is paid-for and stood-behind. We *want* free speech, just not *unmanageable* free speech.

How TIPT Works From the User's Perspective

Every browser screen is divided up into (window) "frames". On a phone, a single frame typically takes up the whole screen. But on bigger screens, there may be lots of frames, each a square or rectangle, which together fill out the browser's window.

Now imagine a new rule for "TIPT-ready" browsers. Each frame MUST have a border in one of four colors with the following meanings for the content within it:

BLACK: Does NOT claim to be in TIPT.

>>> All other colors DO claim to be in TIPT. <<<

```
YELLOW: Browser is busy verifying the claim of being in TIPT.
 GREEN: Has been verified by browser as being in TIPT.
  PINK: Used to be in TIPT, but is no longer.
   RED: Isn't in TIPT, and never was.
```

So, when landing on a new site, the browser frames may all turn yellow, and a few seconds later, all turn green. That would signal that everything on the page is TIPT-registered content. None of it is guaranteed to be *true*, of course, just paid for and stood behind by someone you could find if you wanted to.

But maybe the frames *don't* all turn green; maybe they all stay yellow. That would mean that the browser can't reach TIPT, or TIPT is down.

A pink frame would mean that someone had put the content in TIPT at some point in the past, but it has since expired and has not been renewed.

A red frame would mean it's definitely *not* in TIPT (and is claiming to be), so it's suspicious.

A black frame wouldn't necessarily be suspicious. It could be intentional fiction, fantasy or speculation, or it could require anonymity. Or, possibly—I need to say it—the work of a person too poor to pay to get it *into* TIPT. Everything has its downsides.

But the overall benefit of TIPT to the user should be clear. If she chooses to, she can easily keep herself within the TIPT subset by sticking to green only, or to green and yellow if TIPT appears to be down, or to green and pink if she cares about history.

The Service

TIPT *the service* will be as transparent as we can make it. It will be like one of those see-through fish whose organs can be clearly viewed. TIPT will be completely transparent all the way to its bottom. Because if it isn't, it's worthless. It has to be *mathematically* trusted *at runtime*, similar to the way a blockchain is (but a zillion times faster and cheaper to run—we'll figure it out!!!). In addition, the man on the street will be given access to an interface that lets him crawl *inside* TIPT and watch individual digital artifacts working their way through the system. HUGE payments will be offered (and paid) for any bugs found in the system after its (each) release. Felonies with mandatory prison time will be charged and prosecuted for those who *plant* bugs. I hope you can see that this is not just open *source*, this is open *runtime*.

TIPT Will Store Hashes

What gets stored in TIPT will not be the digital objects themselves, such as full videos, but rather digital *hashes* of them. A hash is a very short, but unique, id that can be calculated for any digital artifact from a specific formula.

Let's say you have a digital photograph. You can run all the (ordered) bits of that photograph through the formula, and you'll get a number out. Some number between 0 and 2^512 let's say. 2^512 is a very big number, for sure, but only if you have to count your way up to it; for our purposes here all you need to do is store some single number in that range. And for that you only need 512 bits, which is tiny. And here's the beauty of hashes, if it's not already clear: regardless of the size of your digital artifact—be it a single sentence or the entire Library of Congress concatenated into a single document—you can always get a(n effectively) unique id out of the formula that's just 512 bits long.

Registering an Artifact

Let's say you want to register your photo with TIPT so that when it gets viewed in someone's TIPT-enabled browser, its border turns green. First, you use the formula to calculate the hash. Then together with that hash, you bundle your user id and the artifact itself (or a pointer to a place where it's stored online), and send it all to TIPT.

```
REQUEST: Register
user_id: 123456
artifact_location: (blank means it's attached)
```

```
hash:5cb124046f7030b9492c45238e2638ea4ac6136f5f553a7603de7c8e67f1df00
75945be492f21c1fb615915726ed988cc667d3dc69bf3cc2992e84c416ea34cc
extend_if_already_registered: true

RESPONSE: Register
status: SUCCESS
extended: false
expiration: May 31, 2025 00:00:00
```

If TIPT calculates the same hash you did, (it's using the same formula), it returns a SUCCESS response, and the hash of your photo is in TIPT.

What Registering Accomplishes

Now, after successful registration, when someone visits your web site with their TIPT-enabled browser, your web site will send down both the photo *and its hash* to the browser. The browser will then do the following.

1. It will display your photo inside a frame with a yellow border.
2. It will calculate the hash of the photo to make sure it matches the hash your web site sent along *with* the photo. If there isn't a match, it sets the border around your photo to red, and it's done.
3. Otherwise, while keeping the border yellow, it will send the hash off to TIPT, asking for verification that the hash has been registered.
4. If TIPT responds "affirmative", it changes the border to green; if "negatory", then red; if "not anymore", then pink.
5. If TIPT fails to respond, the border stays yellow.

If TIPT *does* respond, its response will also contain your user id and real world name, so that the browser can display that information, too, if asked for it.

```
REQUEST: Verify
user_id: 123456
hash:5cb124046f7030b9492c45238e2638ea4ac6136f5f553a7603de7c8e67f1df00
75945be492f21c1fb615915726ed988cc667d3dc69bf3cc2992e84c416ea34cc

RESPONSE:
status: SUCCESS
hash:5cb124046f7030b9492c45238e2638ea4ac6136f5f553a7603de7c8e67f1df00
75945be492f21c1fb615915726ed988cc667d3dc69bf3cc2992e84c416ea34cc
user_id: 123456
user_name: Your Real Name
expiration: May 31, 2025 00:00:00
```

Considerations

That's the picture I wanted to paint. Massively incomplete, but I believe it gets across the general idea for anyone interested in building such a thing. That being said, there are lots of fascinating topics to consider when working out the details, and I want to now randomly visit a bunch of them.

TIPT is a Digital Carbon Dating Implementation

We're living in digital pre-history. No digital artifact from our time can or will be trusted 1000 years from now. How will they definitively know any "newly discovered" artifact from our time is 1000 years old and not 1000 seconds old? They won't.

What would our own paleontology be without dates *built into* the fossils? Would we even believe in evolution? Probably not as strongly.

With TIPT in place, each registered artifact will have an associated and trusted *creation* timestamp. The future digital paleontologist could calculate the hash of any artifact and check for its existence/birthdate in TIPT. This changes everything for those in the future, *and* somewhat for us in the present, too—especially legally.

Trusted User Identities

Part of TIPT, or maybe sister to TIPT, will be a service that maps unique persons to unique numbers, much the way tax systems do. It will help to formalize what banks already do with state-issued ids and utility bills.

But TIPT identities will be *public* information, which spins off a deep and cloudy subtopic on the role of human privacy that I don't want to pursue here. But in general, my take is that privacy is on the permanent decline. Societies that flourish and survive have been, and always will be, those that do a better job at *reducing the need for privacy*.

After all, privacy is unshared information. And while unshared information certainly has its place, the less of it we have the better—because cooperation requires *shared* information. And cooperation wins wars.

Payments

Every interaction with TIPT, other than asking whether or not a given hash is *in* TIPT, will have an associated charge, maybe a charge that fluctuates due to load on the system. Even requests that fail due to user error, like mismatching hashes on a register attempt, will cost the user something.

I suspect the user will be required to have online funds that TIPT can directly decrement. Successful payment will precede all changes to TIPT.

Hashing Penalties

Having browsers calculate hashes over most/all of the data they display will both slow down the user experience and consume a ton of energy. I have no idea how much, but my sense is that if this is the price for having shared, trusted information, we will pay that price. We'll have to. We can't survive in large societies without trust.

Transparency Again

I can't overstate how important transparency in TIPT will be. For example, there will be admin interfaces to query, tweak and fix the system. But none of them can be used behind a curtain. All interaction with the system must not only be audited, (and paid for,) but preserved for all time. The auditing of TIPT must be so complete that TIPT can be run in reverse, like the universe itself; or at least like the math that describes the universe.

Semantics Free

I expect TIPT to be free of artifact *meaning*. For example, when registering your artifact, there will be no description field available for describing what it is you're registering ("a pic of my brother showing how big the fish he didn't catch was"). TIPT won't allow that description field because it can be filled with disinformation.

TIPT will be limited to the simple job of managing hashes and the identities of those that registered them and when. It won't contain the artifacts themselves or descriptions of them.

This entails that if all copies of an artifact are lost, its hash in TIPT has little value. Which is fine. TIPT is not trying to preserve artifacts, it's trying to preserve the registrants and creation dates of the artifacts that do get preserved, however they manage that.

Registration Will Not Entail Ownership

Registration of an artifact is a vote of confidence in its (truth) value, not a claim of ownership. It's a five-star review in modern terminology.

However, creators of artifacts will want to register their own artifacts before others do as evidence of likely authorship.

Artifact Lifetimes

I suspect that artifacts will have lifetimes in TIPT, and that those lifetimes will be extensible via re-registration. So, for example, let's say someone initially registers a very popular artifact, like a particular digital version of the King James Bible. TIPT will then set the expiration date to (let's say) one year from registration date. If another user comes along and re-registers it, that user's act of re-registration will push the lifetime out another year. Very popular artifacts may get "up-voted" like this to ridiculously long lifetimes as a way of demonstrating their importance. TIPT-enabled browsers could surface this information when displaying the artifact.

Artifacts that expire remain in the system and may be renewed at any time.

I don't think deletion will be allowed, but maybe retraction will be.

Long Form

Some digital artifacts, like films, may be very large, and may therefore require many seconds for a hash calculation. This can be worked around with a max artifact size and each segment of the film having its own hash.

Snippets of a video that do not align with the original segmentation of the video would have to be independently registered.

Could the hash of the snippet be mapped within TIPT to the full segment(s) it represents? In other words, could the snippet make a claim to parenthood, and could that claim be verified by TIPT? Maybe?

Pseudo-TIPT Emergence

Let's say I'm right—that we do end up with a TIPT someday. But let's further assume that nobody but me reads what I'm writing here before that happens. In that case, I say that TIPT eventually comes about due to an earlier, *natural* emergence of a poor man's TIPT from our current Internet. Once the presence of this poor man's TIPT is noticed and its significance grokked, it will be obvious how to improve on it in a way similar to what I've described here.

Here's how this will happen. Eventually, good-faith information will be out-competed by disinformation, which is best conceptualized as an invasive species. As disinformation gradually elbows out good-faith information, trust in *all* information (and in one other, for that matter) will gradually decrease. That will create an opportunity for another species of information to rush in and fill the trusted-information void: advertisements. When all information starts to sink in the ocean of trust, information whose publication *has been paid for* will float.

In other words, the implementation of TIPT will begin in a time when everyone is already *voluntarily* paying to publish on the Internet. It will still be free to self-publish on the Internet but no one will do it because self-published artifacts won't find an audience, human or otherwise. If I'm right, we'll reach a time when all content is posted through an *advertiser.* (The meaning of that word may change, or the advertisers may rebrand themselves. Same difference.)

Last Thoughts

Most people I've tried explaining this to can't quite get it. They think, if TIPT doesn't keep out the lies, then what good is it? They're missing the main point, of course, which is: it's *impossible* to keep out the lies in a way that promotes healthy free speech, and healthy free speech is not something we can afford to surrender. Healthy free speech undergirds the whole shebang. The best we can do is to put up some hurdles directly in the path of disinformation. We can charge a fee and we can take names. We can put bouncers at the door.

But if we choose not to, then someday soon you're gonna find yourself hanging on to a steadily deflating life preserver (covered in ads, of course), weeping your regrets into an ocean of disinformation, and finally admitting to yourself, "We shoulda listened to that damn Slippin Fall! (Sniff, sniff.)"Most people I've tried explaining this to can't quite get it. They think, if TIPT doesn't keep out the lies, then what good is it? They're missing the main point, of course, which is: it's impossible to keep out the lies in a way that promotes healthy free speech, and healthy free speech is not something we can afford to surrender. Healthy free speech undergirds the whole shebang. The best we can do is to put up some hurdles directly in the path of disinformation. We can charge a fee and we can take names. We can put bouncers at the door.

But if we choose not to, then someday soon you're gonna find yourself hanging on to a steadily deflating life preserver (covered in ads, of course), weeping your regrets into an ocean of disinformation, and finally admitting to yourself, "We shoulda listened to that damn Slippin Fall."